ISSN: 0974-5823 Vol. 7 No. 1 January, 2022

International Journal of Mechanical Engineering

Grasp Planning for Underactuated Three-Fingers Robot Gripper Using Deep Convolutional Neural Network

Nadin Geies, M.E.H. Eltaib, Hany A. Mohamed and Mahmoud Abdelrahim

Abstract -

Stable grasping of objects is a crucial task for robot manipulators in many industrial applications, in order to implement the desired tasks properly and safely. In this paper, we investigate the use of a suitable mechatronic design for a versatile robot grippers, and the use of deep learning that can be able to successfully grasp various types and sizes of objects in an autonomous fashion. Successful grasping was measured by two terms namely preventing both slippage and deformations of objects being handled. We employed machine vision and deep learning techniques to train the manipulator to decide appropriate contact points with objects to achieve secure and stable grasping. The fabricated gripper is tested by being mounted with the four degrees of freedom SCARA robot in the laboratory. The overall system resulted in a high successful rate of stable grasping for objects with different types, materials and sizes.

Keywords -

Robot gripper, Stable grasping; Convolutional neural networks; Cornell Dataset

1 Introduction

The increasing popularity of robot manipulators in different domains of applications has lead to intensive researchworkinthisarea. Examplesofapplicationsinclude: painting, assembling of parts, welding and assistance in medical operations, to name a few[1][2][3]. In many of such applications, the manipulator end effector plays an essential role in the successful implementation of desired tasks. For instance, in order for the manipulator to pick and place an object from point to point, it is very important to guarantee stable grasping of the object being moved in order to accomplish the task correctly and safely. Therefore, stable grasping of objects attracted the attention of several research work of the literature [4][5][6].

To ensure stable grasping of objects, some researchers proposed to create a 3D model of the object to identify optimal grasping points for the gripper [7][8]. For instance, [9] focused on creating 3D model of each object using 3D simulations to find a proper grasping. However, this technique requires significant labor time to establish big database for matching 3D object models. In addition, to apply this method, physical information like friction and contact model of each object and creating constraints to the object being grasped need to be known a prior, which may not be available in different situations.

To deal with such cases, recent works of the literature proposed to make the manipulator learn from sensors feedback information to be able to handle the grasping of novel objects [10]. These techniques employ tools from the domains of artificial intelligence and machine learning. For instance, the result of [11] [12] depend on hand collected Data set such as Cornell Grasp Dataset [13] and JACQUARD Data set [14] to perform convenient supervised learning. In [15], for the detection of optimal grasping points of the object, several 2D plane images of the object from different angles are required. Note that this method only allows to determine grasping points, but not the required opening angle of the gripper in order to pick the object correctly. To overcome this issue, [16] proposed a deep network for detecting robotic grasps using seven-dimensional representation. To be more specific, grasping detection is divided into two stages, first getting a set of optimal rectangles, then finding the best rectangle for a single successful object grasp achieving an accuracy of nearly 73.9% . The authors of [17] have developed a simplification for the seven-dimensional representation with a five dimensional one including the orientation of the rectangle, the Cartesian coordinates of the center, the width and the height of grasp as shown in Figure 1 .

The authors of [18] used a network performing single-stage regression to graspable bounding boxes his introduced work achieved a higher accuracy and

an error loss. In [19], the technique of Convolutional Neural Network (CNN) was employed for predicting a multi-fingered gripper grasp configuration, where the success probability was maximized to an accuracy of 75.6% for novel objects. The result in [20] worked on grasp detection from RGB-D senor depth images predicting grasp coordinates directly from a network and the network achieved 88 % accuracy.

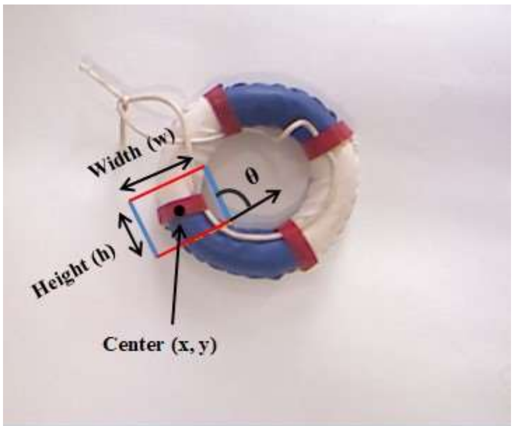

Fig.1.Fivedimensionalrepresentationfortheobject being grasped [17].

Although intensive research work has been devoted, robotic grasp detection ability still very challenging to deal with novel object and to predict optimal gripping points. In this paper we present a deep convolutional neural network approach for grasping unknown objects with a three-fingered under actuated gripper using the rectangle data-driven model on Cornell data-set. the problem of finding the best grasp was addressed by giving a single RGB-D capture of the novel object using a transfer learning approach on correlating to any robotic stable grasping application to improve the previous approaches results and we evaluated the approach on SCARA robot to prove the robustness of our approach. The three fingers under actuated gripper guided by our CNN approach was able to significantlydecreasetheroot-mean-square-error(RMSE) of the grasping trials.

The remainder of the paper is organized as follows. In Section 2, we present the hardware setup used in this work. The grasping technique is provided in Section 3. The obtained results are given in Section 4. In Section 5, we discuss the conclusion and future work directions.

2 Setup Description

For the validation of our approach through this application, several hardware components were involved that will be discussed in this section. As in any deep machine learning application, a PC with a high computational power is required for the neural network training to work competitively.

In addition to the SACRA robot, some information of control infrastructure will be added and described. The key component of this research is the modified three finger under-actuated gripper [21] attached to the SCARA robot. A Microsoft Kinect v1 for image acquisition was in interface with the PC work station. In addition to that, the methodology of our applied algorithm is discussed mentioning the Deep convolutional network approach architecture. As well as presenting the pre-compiled dataset and our collected data.

2.1 Underactuated Gripper Design

For the robot gripper design, we had numerous options for carrying out our stable pick and place application work.The wide range of grippers basically differ according to, number of fingers, degrees of freedom (DOF), actuation method used and applications. We shortened our choices to two different designs of three fingers undeactuated grippers. The two proposed designs are considered to be three-Finger adaptive grippers that differs only in the fingers adjustments and the actuation mechanism architecture. In both two designs, the Gripper is actuated by single servo motor that is able to carry a relatively large payloadcomparedtothegripperweight. Eachfingerofthe robotic hand has 2 DOF (degrees of freedom), one degreeof-actuation and the other degree of freedom earned from the spring (passive element) between the two phalanges for extension and flexion of the finger. In the first design shown in Figure 2, the three fingers have independent angular motions.

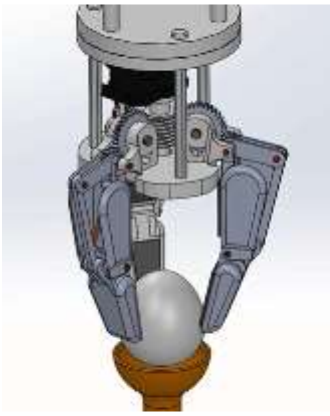

Fig. 2. First proposed design of underactuated gripper.

The second proposed design shown in Figure 3 consists of two adjacent fingers moving simultaneously while the third finger moves independently.

Fig. 3. Second proposed design of underactuated gripper.

To assess the efficiency of both designs, we experimentally carried out numerous trials with diverse objects, which were classified into three geometrical categories (spherical, planar and cylindrical) as seen in Figure 4.

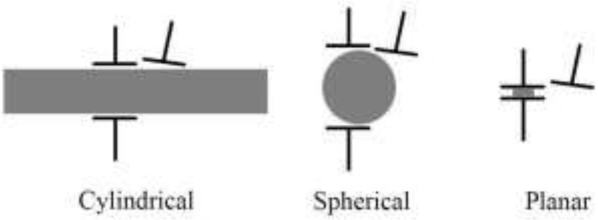

Fig. 4. Gripper configurations for grasping different objects [22]

There were some limitation for the first design to deal with objects with a planar geometry unless it had a small volume within a specific range. Therefore, we chose the second design with the breakaway clutch mechanism architecture to investigate the presented deep learning algorithm. The design mainly consists of a helical gear mounted on the motor transmitting motion to the three helical worm gears connected to each finger.

By analyzing the forces acted on the helical gear there is an axial thrust force F_x that works in tangential plane tending to push the driven gear along the carrying shaft

$$Fx = Ft \tan\Psi \qquad (1)$$

where F_t is the transmitted force which equals:-

$$Ft = 2Ta * D \qquad (2)$$

ψ is a helix angle D is the pitch diameter of the gears

and Ta is the transmitted actuated torque to the fingers .

2.2 SCARA Manipulator

The SCARA Robot (Selective Compliant Assembly research Robot), has been selected to confirm the robotic grasping ability of the trained CNN model. The SCARA robot is Traditionally a 4-axis with 4 DOF.

Fig. 5. SCARA robot attached with the three finger under actuated gripper.

The robot arm is moving to any X-Y-Z coordinate point in their work-space and a fourth axis of motion which is the wrist rotate (Theta). The complete robotic system is controlled through a LABVIEW program providing an inverse kinematic solver. This solver controls the SCARA robot joint motions to pick up the different objects with thepredictedgraspconfigurationpoints. ThesystemMID7654 controller is ideal for any industrial and laboratory application . It connects the Robot motors, encoders, limit switches and other motion hardware to National Instruments motion controllers.

2.3 Kinect Sensor

A Microsoft Kinect xbox v1 which is a web-cam style sensor is used in many computer vision algorithms and applications, mounted on top of the work space. It is used to capture both colour, point clouds and depth (RGB-D) images of the work area used for object tracking, recognition, detection and tracking .The resolution of the sensor camera is 640×480.

3 Grasp Planning Technique

In this section, we explain the proposed method to achieve successful grasping of objects.

3.1 Pre-compiled data

From the previous work, it was seen that deep learning needs a large volume of training data creating large data-sets with manually labeled images for robotic grasping applications. This process is considered to be hard and time consuming, as it requires months of lab work. In addition to that numerous data and simulation iteration must be held to get proper results. That's why Pretraining was proposed [23], to use the weights of the CNN model that was trained on a large dataset like AlexNet, ResNet-50 etc.., to be able to transfer the learning ability to a smaller data-set. We trained our model on the Cornell grasp dataset, which was provided by Robot Learning Lab[13].The CGD is considered to be one of the most important grasping data-set that most of the transfer learning approaches were evaluated by in robotic stable grasping researches. It is also well known with having diverse objects as seen in figure 6 for any general grasp application with good quality and adaptability . The CGD contains

Fig. 6. Sample of Cornell Grasping Dataset [13].

grasp rectangles for about 280 different objects, containing more than one image and point clouds for each object with different poses and orientations. The CGD includes both valid and invalid grasp rectangles.

The rectangle metric defined a successful grasp evaluating method in grasp detection on the Cornell Grasp Dataset[18]. The rectangle metric cost function that evaluates the grasp stability of our approach depends first on the distance between both the predicted grasp and the labeled ground truth presented by the Jacquard index to be less than 25%. The Jacquard index is given by Equation 3.1.

$$J(Grasp_{pred}, Grasp_{true}) = \frac{J(Grasp_{pred} \cap Grasp_{tru}}{J(Grasp_{pred} \cup Grasp_{tru})}$$

Second the difference between the predicted grasp angle and the ground truth grasp angle is less than 30 degrees.

3.2 Architecture of the Convolutional Neural Network

The approach mentioned in the paper is a deep convolutional neural network. By mentioning convolutional network, we mean it is an example of a simple filter to an input that might be any data with a grid pattern, mainly an image or could be multiple images. This filter extracts features and patterns such as edges through back-propagation by using multiple network blocks, such as convolution layers and pooling layers and fully connected layers to create a feature map to the output.Theoretically, for any DCNN the increased number of convolution layers should lead to a better performance but it could also be hard to be optimized which leads to overfitting and a high training error. That’s why for our transfer learning approach as seen in figure 7, where we used ResNet-50, which contains a fifty

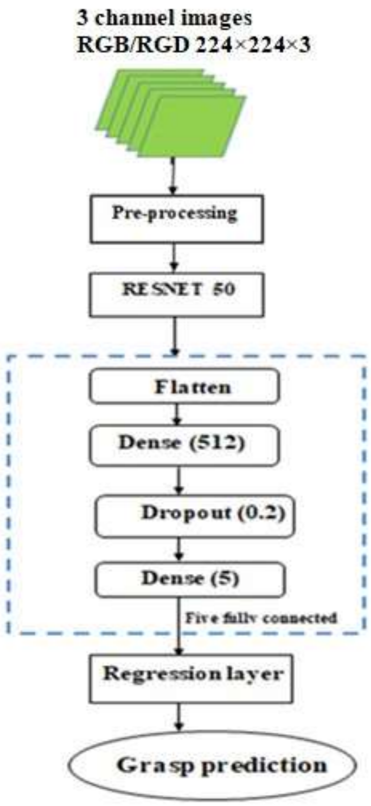

Fig. 7. The complete CNN network architecture.

layer deep residual model, for solving the grasp configuration detection problem.

TheResNet-50isconsideredtobeasimplefeed-forward CNN that first introduced the residual concept by bypassing few layers to minimize training times and reducing the error rate.

The output from the was then ResNet-50 to a fullyconnected layer that contains 512 units, and it is followed by a 0.5 drop-out factor. A rectifier linear unit (RELU) is used for activation. Finally we were able to recognize the object and predict the best grasping rectangle.

3.3 Proposed Grasping System

The final grasping mechanism, starts first with the Microsoft Kinect v1 capturing both coloured RGB images and point cloud data of each object of our dataset. Our dataset consist of various objects for instance, (the paper towel, the orange, the sunglasses,the pipe and the colored ring), which we plan to detect its convenient grasping rectangle for the stability of grasping objects.

Then the blue channel of the RGB image captured is replaced with depth information to create coloured depth (RG-D) images. The processed RG-D images of the objects will then be sent through the trained deep convolutional neural network (DCNN) to be able to estimate the best grasping rectangle presented by (x, y, θ, h, w) parameters as a predicted network output as seen in figure8.

Fig. 8. The different objects used in the experimental work with some grasp predictions showing valid grasp predictions.

These parameters represent an oriented rectangle on the image plane with a centre point at (x, y), a height (h), and a width (w). The grasp detected parameters are then sent from MATLAB to the labview program which contains inverse kinematic solver to map them according to the SCARA robot work space for object grasping. After the object has been successfully grasped and placed at the fall-off zone, the robotic arm will be reset for the next object as seen in figure 9. For our grasping experiments, we chose a diverse set of 6 objects, the weight of the objects is at most 2 kg with size within 0.4 m x 0.4 m . Some of the choosen objects were not part of the training dataset and were absolutely new to the grasp detection algorithm. Each of the objects will have about 20 trials as preformedinmostofthepreviouswork[20]. Thetrialswas classified into two categories (stable and unstable grasp). A comparison between the experimental and theoretical results was then held to choose either to extend the trials for another 20 trial or not.

4 Results

The training model was held in MATLAB platform with the use of the Neural Networks toolbox [24]. This toolbox contains several functions to build CNN layers for any classification or regression application. In addition to the toolbox ability of transfer learning with several pre trained models as the pre-trained ResNet-50 model, we used in our model.

The network end to end model was compiled with the stochastic gradient descent optimiser with an initial learning rate of 0.001 and trained for just 8 hours, as the used pre-trained model saved much time. We trained the network until the RMSE reached its steady state and the error was decreased by 91.25% as seen in figure 10.

by evaluating the trained CNN on the CGD, we used about 100 out of 885 for validation and 600 for the main training data. The final predicted rectangles for the objects afterproceedingthelearningapproachmadetherobotable to attempt a stable grasp. Our algorithm achieved some worthy successes as it was able to detect and accomplish grasps for a paper roll, glasses, Orange, kitchen scoop and a colored plastic shaped ring. In addition to an irregular pipe which is considered totally unsimilar to any object in the dataset. In figure 11 we can see the SCARA robot during the grasping attempt trials for the previously introduced 6 objects. The three finger under actuated gripper attached to the SCARA robot was differently adjusted for each object following the network model output configuration. Finally, the SCARA robot was able to grasp most of the object in a stable manner within a good accuracy percentage. In table 1, the results of our robotic experiments for 6 objects is presented, a total of 20 trials for each object. Using our algorithm, the SCARA was able to successfully accomplish a stable grasp in 80% of the trials. In 10% of the trials, our algorithm detected a valid graspthatwasnotexecutedcorrectlybySCARAandinthe rest of trails our network wasn't able to detect valid grasp rectangles. In table 2, a comparison was held between our system accuracy and the previous work accuracy with showing a modest success rate. The experimental results

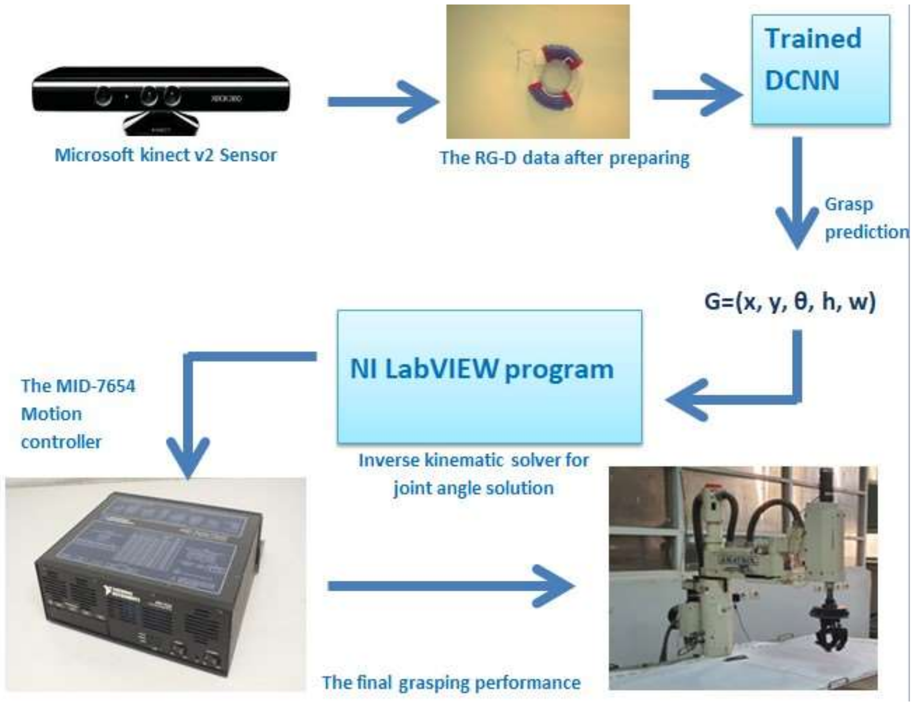

Fig. 9. The final robotic grasp detection process based on our deep learning approach

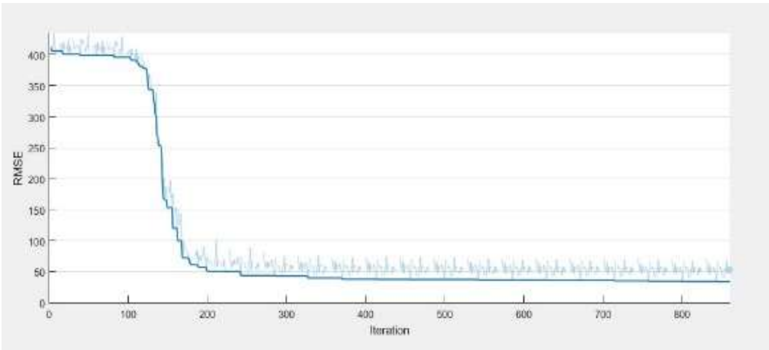

Table 1.
first Table
1. first

Fig. 10. Network training plot of the RSME , that decreases with the number of iterations

Table 1. first Table 1. first experimental results

Object	stable trial	unstable trial	Accuracy
colored ring	14	6	70
orange	19	1	95
pipe	17	3	85
glasses	16	4	80
Towel	18	2	90
kitchen scoop	12	8	60
Average			100
Overall			80

were convenient so there was no need for exerting more experimental trials.

Authors	Algorithms	Success %
Lenz et al. [17]	Sliding window	73.9
Jiang et al. [25]	Fast search	60.5
Watson et al. [26]	Direct regression, AlexNet	78.00
shehan caldera [20]	Vector regression, (RGB)	78.8
Our result	Vector regression (RGB)	80

5 Conclusion

To conclude our work we aimed to solve the problem of unsteadiness of grasps held by a SCARA robot, by trying to mimic human being grasping abilities. First a three finTable 2. Comparison between different transfer learning techniques in grasp detection.These tests were performed on the Cornell Grasp Dataset

ger under-actuated gripper was drawn on solid works and 3D printed to be attached to the robot to investigate the use of our network model. The gripper was able to hold relatively high payload in comparison to the simple design structureandcost. Inadditiontopresentingaroboticgrasp detection system, that predicts the best grasping rectangle configuration for the robotic three finger end effector. To finally be able to hold novel object in a stable manner. Features were extracted from both RGB and RGB-D images through the five fully connected layers and the regression layer of our Deep Convolutional Neural network reaching the grasp predicted configuration. A real time implementation of Our model were carried out with 6 novel objects achieving an accuracy of more tan 80%.

Fig. 11. Stable grasp of six different objects using 3-finger underactuated gripper attached to SCARA robot

References

[1] A. Khurshid, Abdul Ghafoor, and Afzaal Malik. *Robotic Grasping and Fine Manipulation Using Soft Fingertip*. 08 2011. ISBN 978-953-307-373-6. doi:10.5772/23697.

[2] Kevin Tai, Abdul-Rahman El-Sayed, Mohammadali Shahriari, Mohammad Biglarbegian, and Shohel Mahmud. State of the art robotic grippers and applications. *Robotics*,5(2),2016. ISSN2218-6581. URL https://www.mdpi.com/2218-6581/5/2/11.

[3] Lionel Birglen and Thomas Schlicht. A statistical reviewofindustrialroboticgrippers,2018. ISSN07365845. URL https://www.sciencedirect.com/ science/article/pii/S0736584516304240.

[4] Dong-Eon Kim, Ailing li, and Jang-Myung Lee. Stable grasping of robotic hand using objective function. pages 538–543, 07 2019. doi:10.1109/AIM.2019.8868615.

[5] Hao Dang and Peter K. Allen. Learning grasp stability. In *2012 IEEE International Conference on Robotics and Automation*, pages 2392–2397, 2012. doi:10.1109/ICRA.2012.6224754.

[6] H. Dang and P. Allen. Stable grasping under pose uncertaintyusingtactilefeedback. *AutonomousRobots*, 36:309–330, 2014.

[7] Angel Valencia, Roger M. Idrovo, Angel Sappa, Douglas Plaza Guingla, and Daniel Ochoa. A 3d vision based approach for optimal grasp of vacuum grippers. pages 1–6, 05 2017. doi:10.1109/ECMSM.2017.7945886.

[8] Brayan Zapata-Impata, Carlos Mateo, Pablo Gil, and Jorge Pomares. Using geometry to detect grasping points on 3d unknown point cloud. pages 154–161, 06 2017. doi:10.5220/0006470701540161.

[9] Y. Li, Y. Yu, and S. Tsujio. An analytical grasp planning on given object with multifingered hand. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, volume 4, pages 3749–3754 vol.4, 2002. doi:10.1109/ROBOT.2002.1014296.

[10] Chao Wang, Xuehe Zhang, Xizhe Zang, Yubin Liu, Guanwen Ding, Wenxin Yin, and Jie Zhao. Feature sensing and robotic grasping of objects with uncertain information: A review. *Sensors*, 20:3707, 07 2020. doi:10.3390/s20133707.

[11] You Li, Qujiang Lei, ChaoPeng Cheng, Gong Zhang, Weijun Wang, and Zheng Xu. A review: machine learning on robotic grasping. page 54, 03 2019. doi:10.1117/12.2522945.

[12] Jiahao Zhang, Miao Li, Ying Feng, and C. Yang. Robotic grasp detection based on image processing and random forest. *Multimedia Tools and Applications*, 79:2427–2446, 2019.

[13] Cornell University. *"Robot Learning Lab: Learning to Grasp*. URL http://pr.cs.cornell.edu/ grasping/rect.

[14] Amaury Depierre, Emmanuel Dellandréa, and Liming Chen. Jacquard: A large scale dataset for robotic grasp detection. *CoRR*, abs/1803.11469, 2018. URL http://arxiv.org/abs/1803.11469.

[15] Fangyi Zhang, Jürgen Leitner, Michael Milford, Ben Upcroft, andPeterCorke. Towardsvision-baseddeep reinforcement learning for robotic motion control, 2015.

[16] Yun Jiang, Stephen Moseson, and Ashutosh Saxena. Efficient grasping from rgbd images: Learning using a new rectangle representation. In *2011 IEEE International Conference on Robotics and Automation*, pages 3304–3311, 2011. doi:10.1109/ICRA.2011.5980145.

[17] Ian Lenz, Honglak Lee, and Ashutosh Saxena. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research*, 34(4-5):705–724, 2015. doi:10.1177/0278364914549607. URL https://doi.org/10.1177/0278364914549607.

[18] Lu Chen, Panfeng Huang, and Zhongjie Meng. Convolutional multi-grasp detection using grasp path for rgbd images. *Robotics and Autonomous Systems*, 113, 03 2019. doi:10.1016/j.robot.2019.01.009.

[19] Q. ; Chenna Lu, K. ; Sundaralingam, B. ; Hermans, andT. Planningmulti-fingeredgraspsasprobabilistic inference in a learned deep network. *arXiv 2017*, page 03289., 1804.

[20] S. Caldera. *Learning to grasp in unstructured environments with deep convolutional neural networks using a Baxter Research Robot*. 2019.

[21] N. A. Geies, M. Abdelrahim, M. ElTaib, and H. A. Mohamed. Grasping stability analysis of an underactuated three finger adaptive gripper on matlab simmechanics. In *2020 16th International Computer Engineering Conference (ICENCO)*, pages 141–146, 2020. doi:10.1109/ICENCO49778.2020.9357396.

[22] Kuat Telegenov, Yedige Tlegenov, Shahid Hussain, , and Almas Shintemirov. Preliminary design of a three-finger underactuated adaptive end effector with a breakaway clutch mechanism. *Journal of Robotics and Mechatronics*, 27(5):496–503, 2015. doi:10.20965/jrm.2015.p0496.

[23] ZhiyuanTang, DongWang, YiqiaoPan, andZhiyong Zhang. Knowledge transfer pre-training, 2015.

[24] Mathworks. *Deep Learning in Matlab*. Available online:

https://au.mathworks.com/help/nnet/ug/deeplearningin-matlab.html (accessed on 28, April 2017.

[25] Y. Jiang, S. Moseson, and A. Saxena. *"Efficient grasping from rgbd images: Learning using a new rectangle representation, " in 2011 IEEE International Conference on Robotics and Automation*. May 2011.

[26] J. Watson, J. Hughes, and F. Iida. *Real-world, realtime robotic grasping with convolutional neural networks, " in Towards Autonomous Robotic Systems*. Cham: Springer International Publishing, 2017.

